

1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN  
12400 Wilshire Blvd., 7th Floor  
Los Angeles, California 90025-1026  
(310) 207-3800

## BACKGROUND

### (1) Field of the Invention

The invention relates to network communications. More specifically,  
5 the invention relates to traffic management in a network.

### (2) Background

Downstream internet traffic flows on oversubscribed copper lines at  
rates DS-1 and below dominate the performance attributes of internet  
applications. Large carriers have been deploying frame relay access switches  
10 since the early nineties. ILECs and CLECs have deployed large footprints of  
first generation digital subscriber line access multiplexers (DSLAMs).  
Likewise, Internet service providers (ISP's) and cable operators have a large  
embedded base of legacy routers, hubs and cable modem termination systems  
(CMTS). These deployments have resulted in a large embedded base of legacy  
15 equipment with very limited traffic management features. Typical queuing  
systems are FIFO based and often a FIFO is shared across lines allowing  
customers to interfere with each other. One result of this FIFO queuing is that  
two flows directed to the same line may not be delivered in desirable order.  
For example, a packet or cell of a web page download or e-mail may be  
20 delivered in advance of packet or cell of the next video frame.

Bandwidth demands are continually increasing. This ever-growing  
demand for bandwidth necessitates traffic management techniques. While  
existing "last mile" infrastructure creates a performance bottleneck for  
downstream traffic flows, the cost of replacing this existing legacy equipment  
25 would be very high. It is useful to add traffic management capabilities to the  
network without replacing the legacy equipment.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention is illustrated by way of example and not by way of limitation in the figures of the accompanying drawings in which like  
5 references indicate similar elements. It should be noted that references to "an" or "one" embodiment in this disclosure are not necessarily to the same embodiment, and such references mean at least one.

**Figure 1** is a block diagram of a system of one embodiment of the invention.

10 **Figure 2** is a block diagram of an aggregator and remote physical ports networked thereto in one embodiment of the invention.

**Figure 3** is a generalized flow diagram of traffic management in one embodiment of the invention.

## DETAILED DESCRIPTION

**Figure 1** is a block diagram of a system of one embodiment of the invention. Server nodes 102 and 103 may be any server nodes that might exist on the world wide web. Such server nodes may stream audio, stream video, serve web pages, serve e-mail, or provide other types of data across a distributed network, such as web 100, through an aggregator 104 across a trunk line 110 through a switch 112 to a line 113 and finally to customer premise equipment (CPE) 114. Trunk line 110 may be any broadband communication link, for example, a DS-3 or an OC-3 line. Flows from aggregator 104 through the switch 112 toward CPE 114 are regarded as downstream flows. Typically, downstream flows originate at a server node such as server node 102. Frequently, switch 112 has very limited traffic management capabilities. Aggregator 104 includes a line card 106 having a traffic manager 108 thereon.

The traffic manager 108 implements a model of the physical line rate of the line 113 of the switch 112. The model includes a traffic shaper which limits traffic destined to that port to that line rate. By modeling the physical bit rate of the lines of the switch 112, the traffic manager 108 knows when the incoming traffic from such servers as 102 and 103 at the aggregator destined for a particular line of the switch 112 exceeds the line rate for that line. When it does, even instantaneously, the traffic manager 108 only sends packets or cells to the switch 112 at that line rate, queuing the excess traffic within the traffic manager 108. Within the traffic manager 108, sophisticated traffic management capabilities may be invoked to control the individual flows destined for line 113. For example, video packets received from server 102 may be sent out before an e-mail received from server 103. In addition, only low priority packets may be discarded according to some packet discard policy when queues reach a certain queue threshold.

As long as the legacy switch 112 does not receive traffic for a particular port at a bit rate greater than the port is able to carry, nothing is queued at the FIFO buffers of the legacy switch 112. The traffic manager may insure that the legacy switch 112 only receives a new packet or cell from the aggregator when  
5 the previous packet or cell has already been sent out on the line 113.

Accordingly aggregator 104 becomes the only place where traffic management occurs. The legacy switch 112 becomes transparent to traffic management because traffic to the line 113 is already managed at the upstream aggregator 104 and the legacy switch 112 adds no queuing delay to any packets or cells.

10 The aggregator 104 can thus be said to remotely manage the traffic of the legacy switch 112. In one embodiment, the traffic manager is implemented on an ASIC.

**Figure 2** is a block diagram of an aggregator and remote physical ports networked thereto in one embodiment of the invention. At the network edge  
15 is a legacy switch or demultiplexer 232 which has a plurality of remote physical ports (RPPs) 236. Each such port operates at a particular transfer rate. For example, remote ports may operate at DS-1 rates or below. A FIFO 234 associated with port 236 is provided in the event that the incoming rate on the trunk 230 exceeds the RPPs transfer rate. Subsequent data would typically be  
20 queued in the FIFO. A legacy switch or demultiplexer 232 distributes incoming transmission units from the trunk line 230 to the appropriate RPP. By way of example, the trunk line 230 may be a DS-3 connection which implies it has 28 times the capacity of a DS-1.

A remote logical port (RLP) traffic manager 108 consists of a flow  
25 manager 109 followed by a RLP model 201. Thus, where there are L physical ports, where L is an arbitrarily large positive integer, there will be L RLP traffic managers 200, L flow managers 205 and L RLP models 201, resulting in a one-to-one correspondence. L is expected to be rather large, such that the aggregate bandwidth of the L RPPs is much greater than the capacity of the

trunk. All flows directed to a particular remote physical port are handled by its corresponding RLP traffic manager. The RLP traffic manager 200 is for remote RPP 236. The RPP model 201 may receive N flows 202 of packets or cells, containing such information as streamed video or audio, and, for example, M  
5 flows 204 of packets or cells, containing such information as a web page download or e-mail.

In the subsequent discussion, packets, frames or cells are referred to as transmission units. Illustratively, a transmission unit may be, for example and without limitation, a layer 3 packet which may have a variable length, a  
10 layer 2 frame with a variable length, or an asynchronous transfer mode (ATM) cell which has a fixed length. Illustratively, flows are a sequence of transmission units associated with a particular customer, a particular connection, or a particular application such as video, or a combination of such associations.

15 The function of the flow manager 205 is to provide better bandwidth management of traffic flows than are provided at the legacy switch 232. Better bandwidth management is accomplished by providing more features capable of differentiating the flow characteristics of flows than are available at the legacy switch. The following are illustrative of flow management.  
20 Instead of a shared FIFO queue 234 for all flows, a queue 208 is provided for each incoming flow 202, 204. Transmission unit discard policies 206 may be applied to the buffers of both shaped and unshaped flows 202, 204. An example of a discard policy may be: "if queues containing video information are at least 1/4 full and queues containing e-mail information are at least 2/3  
25 full, discard the last transmission unit containing e-mail information from its queue." Flows containing video or audio streaming information may be advantageously shaped in a flow shaper 210. The flow shaper 210 smoothes the flow of transmission units for reception by a CPE device like a PC. By "shaping," it is meant that the eligibility of a transmission unit for

transmission is determined by the time elapsed since the transmission of the previous transmission unit from that flow. Both shaped and unshaped flows may then scheduled for transmission. For example, a transmission unit in a queue containing video information may be scheduled for transmission ahead of a transmission unit in a queue containing e-mail information. All the flows, both shaped and unshaped, are scheduled by RLP scheduler 214. Shaped flows may be given higher priority than unshaped flows. When thus prioritized, if at any time, the sum of the flow exceeds the RPP rate, e.g. DS-1, unshaped flow will backup in the queues of the traffic manager, while shaped flows are handled on a best efforts basis. The RLP scheduler 214 presents a transmission unit of the most urgent flow to the RLP model.

In one embodiment of the invention, once the individual flows are controlled using some or all of the traffic management techniques described above, the next scheduled transmission unit of a flow is presented to the RLP model to further determine eligibility for transmission. The function of the RLP model is to determine when the transmission unit is eligible to be presented to the trunk scheduler. The RLP model includes an RLP shaper and an RLP model data structure 218. The RLP data structure 218 is loaded with shaping parameters that correspond to the transmission rate,  $r$ , of the RPP. In one embodiment, the RLP shaper 216 assures that the transmission unit is made eligible for trunk scheduling no sooner than after an elapsed time,  $t$ , since the last transmission unit for that RPP was transmitted on the trunk. The elapsed time  $t = s/r$ , where  $s$  = the size of the previous transmission unit (in bits) and  $r$  = the rate of the RPP in bits/second. This is simply the duration it takes to transmit a transmission unit on the RPP.

The parameter  $r$  is a variable obtained by provisioning from the management plane 224. The parameter  $s$  is stored in the database with each transmission unit sent. It may also be constant as in the case of fixed length ATM cells. The trunk scheduler also feeds back a parameter,  $T$ , which is the

time at which the previous transmission unit destined for the RPP was actually transmitted on the trunk. It is also stored in the data structure 218 until the next transmission unit for that RPP is transmitted. It is within the scope and contemplation of the invention to use parameters such as the  
5 inverse of a rate to calculate eligibility time of the transmission unit. The RLP model 201 assures that the RPP will be able to transmit the previous transmission unit out on the line before the next transmission unit arrives.

Each of these RLP model parameters are associated with a RLP in the data structure 218. Any transmission unit destined for a particular RPP is  
10 associated with an RLP. The RLP may be identified from the transmission unit headers by various methods. One method is to assign a unique connection identifier to all traffic destined for a particular RPP. For example an ATM VPI or MPLS label may identify the RLP. The individual flows then may be identified by IP addresses encapsulated within the MPLS packet or  
15 ATM cell. Or in another embodiment, the flows are identified with virtual circuit identifiers (VCIs), while the RLP is identified with a virtual path identifier (VPI). In yet another embodiment, a VCI identifies a flow for processing in the flow manager 205, and a multitude of VCIs identify the RLP. This may be accomplished by looking up a VCI in a lookup table (LUT) (not  
20 shown) to find an associated RLP identifier. Multiple VCIs all going to the same destination RPP will have the same RLP identifier associated with them. A second lookup of the RLP identifier in a second LUT will find the shaping parameters associated with the RLP. These are but a few of the many ways to distinguish flows and RLPs from transmission unit headers.

25 Flow shaper 210 forms its shaping based on flow parameters from flow parameter database 212 as described. The flow parameter database 212 may be populated by a control plane 220. Control plane 220 is basically a connection or flow manager that receives connection or flow policy information from the signaling network or from the management plane. Control plane 220 includes



a connection admission control (CAC) that matches inflows with downstream bandwidth. In one embodiment of the invention, the CAC ignores the RLP structure and merely subtracts the transmission rate of incoming flows from the available outgoing transmission rate of the trunk. This method enables  
5 the trunk 230 to achieve statistical gain. In other words it operates in a work conserving manner at least some or most of the time.

The RLP shaper 216 shapes the scheduled flow established by the RLP scheduler 214. Shaping by the RLP shaper 216 is based on, for example, legacy port rates provided by the RLP model data structure 218. RLP model data  
10 structure 218 may be populated by the management plane 224 as described above. Population of the RLP model database 218 may be by direct entry from a manager via user interface device 226. Alternatively, management protocol, such as simple network management protocol (SNMP) could be used to query port management information buffer (MIB) in the legacy switch for port  
15 information and corresponding transmission rate, sometimes implied by the type of port. For example, if the type of port is DS-1, then it implies a transmission rate of 1.544 Mb/s. Scripts could be used to automate the queries and collect responses, and further, could then be used to automatically populate the RLP model data structure 218.

20 Each RLP model indicates eligibility of its shaped flow to the trunk scheduler 228. In one embodiment, flow is only deemed eligible if sending a transmission unit will not cause a backup in the downstream queue. This can be determined based on the port rate and the timing of a previous transmission as explained above.

25 The trunk scheduler 228 schedules a trunk flow from the set of eligible transmission units of all the remote logical ports. In one embodiment, the transmission units are scheduled in a work conserving way for the trunk. It is expected that relatively few RLP are subject to shaping simultaneously. The other flows can fill up the trunk to make it work-conserving. By statistically

multiplexing, the trunk scheduler 228 is able to supply many more physical ports than the trunk capacity alone would permit. Both levels of shaping and scheduling may be performed using pointer manipulation within the queuing structures that receive the flows.

5           The above described a hierarchical dual level shaping and scheduling system at an upstream traffic manager node that permits flows to be individually shaped and scheduled such that a downstream flow at the port of a legacy switch, router or network provides a quality of service that the port's own traffic management facilities could not guarantee. This allows legacy  
10 equipment to appear to have traffic management capability where it is not present. Accordingly, the capital cost of replacing such equipment to achieve the desired quality of service may be avoided. Remote traffic management could be provided to hundreds or even thousands of RPPs using just one traffic manager which might only be one or a couple of ASICs on an aggregator  
15 line card.

While the discussion above relates to a legacy switch, this is merely illustrative. Particularly, a frame relay switch, ATM switch, Ethernet hub, router, cable modem termination system (CMTS) or even a network of these elements, may be modeled and managed as discussed above. By way of  
20 additional example, for a network of elements having significant trunking capacity, assuming the data bottlenecks in a last line leading out of the network, that line can be modeled as the RPP. It is also within the scope and contemplation of the invention to employ additional levels of shaping and scheduling, particularly where a downstream network is to be modeled.

25           **Figure 3** is a generalized flow diagram of traffic management in one embodiment of the invention. At functional block 302 a transmission unit is received from an incoming trunk and distributed to an appropriate RLP traffic manager. Block 301 corresponds to flow management within a traffic manager. At functional block 304 the transmission unit (TU) is placed in a

queue associated with a particular priority or flow. At functional block 306 TUs are discarded from the queue or queues according to some discard policy that is able to differentiate between the discard rates of at least two flows. A TU is indicated to be eligible for RLP scheduling once it has satisfied some

5 flow-shaping requirements at functional block 308. The particular requirements may be arbitrarily established, and may include any known or subsequently developed flow-shaping techniques.

At functional block 310 the most urgent TU is indicated to the RLP model based on RLP scheduling policy. The RLP flow scheduling policy, like

10 the flow-shaping requirements, may be an arbitrary scheduling policy. Box 311 corresponds to the operation within the RLP model. At functional block 312, the eligibility of the most urgent TU is determined based on shaping the flow to match the RPP transmission rate.

One or more, but not necessarily all of the blocks 304, 306, 308, and 310

15 may be used to provide better traffic management than the legacy switch, depending on the QoS features of the legacy switch.

Box 315 corresponds to operation at the trunk scheduler. At functional block 314, when the TU for the RLP model is eligible, and the most urgent of all the TU's from all the RLP traffic managers based on the established trunk

20 scheduling policy, the TU is transmitted out and the transmission time of the TU is reported back to the RLP model at functional block 316. The trunk scheduling policy may vary in sophistication from one embodiment to the next. For example, in one embodiment the trunk scheduling may be simple, first in first out (FIFO). Nevertheless provided that more sophisticated

25 management is used in the traffic manager, improved quality of service is provided to the RPPs. Alternatively, sophisticated scheduling policies may be implemented by the trunk scheduler in addition to any other policies applied upstream.

In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will, however, be evident that various modifications and changes can be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended  
5 claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

**B**